

WHAT IS CLAIMED:

1. In a computer system, a method of sequencing a nucleic acid sequence, the method comprising the steps of:

inputting hybridization intensities into a computer for a plurality of nucleic acid probes, the nucleic acid probes hybridizing with the nucleic acid sequence under conditions that do not allow identification of all nucleic acid probes that are perfectly complementary to part of the nucleic acid sequence; and

sequencing the nucleic acid sequence according to selected nucleic acid probes.

2. The method of claim 1, further comprising the step of selecting nucleic acid probes utilizing mismatch information among the nucleic acid probes.

3. The method of claim 1, further comprising the step of selecting nucleic acid probes with the highest numbers of single base mismatch neighbors among the probes, a single base mismatch neighbor being another probe that has the same sequence except for a single base that is different.

4. The method of claim 1, further comprising the step of selecting nucleic acid probes that have fewer than a predetermined number of base mismatches with another probe.

5. The method of claim 1, further comprising the step of selecting nucleic acid probes with highest hybridization intensities.

6. The method of claim 5, wherein the nucleic acid probes with the highest hybridization intensities are above an intensity threshold.

7. The method of claim 1, wherein the sequencing step includes the steps of:

aligning the selected nucleic acid probes so that bases that are common overlap; and

sequencing the nucleic acid sequence according to bases that occur most often at each position in the aligned nucleic acid probes.

8. The method of claim 7, wherein the nucleic acid sequence is sequenced as complementary to the bases that occur most often.

9. A computer program product that sequences a nucleic acid sequence, comprising:

computer readable code that receives as input hybridization intensities for a plurality of nucleic acid probes, the nucleic acid probes hybridizing with the nucleic acid sequence under conditions that do not allow identification of all nucleic acid probes that are perfectly complementary to part of the nucleic acid sequence; and

computer readable code that sequences the nucleic acid sequence according to selected nucleic acid probes;

wherein the computer readable code are stored on a tangible medium.

10. In a computer system, a method of sequencing a nucleic acid sequence, the method comprising the steps of:

inputting hybridization intensities for a plurality of nucleic acid probes;

selecting nucleic acid probes with highest numbers of single base mismatch neighbors among the probes, a single base mismatch neighbor being another probe that has the same sequence except for a single base that is different; and

sequencing the nucleic acid sequence according to the selected nucleic acid probes.

11. The method of claim 10, further comprising the step of selecting nucleic acid probes with highest hybridization intensities.

12. The method of claim 11, wherein the nucleic acid probes with the highest hybridization intensities are above an intensity threshold.

13. The method of claim 10, wherein the step of selecting includes the step of calculating a number of single base mismatch neighbors for each nucleic acid probe.

14. The method of claim 10, wherein the step of selecting includes the step of comparing double base mismatch neighbors, a double mismatch neighbor being another probe that has the same sequence except for two bases that are different.

15. The method of claim 10, wherein the sequencing step includes the steps of:

aligning the selected nucleic acid probes so that bases that are common overlap; and

sequencing the nucleic acid sequence according to bases that occur most often at each position in the aligned nucleic acid probes.

16. The method of claim 15, wherein the nucleic acid sequence is sequenced as complementary to the bases that occur most often.

17. A computer program product that sequences a nucleic acid sequence, comprising:

computer readable code that receives as input hybridization intensities for a plurality of nucleic acid probes;

computer readable code that selects nucleic acid probes with highest numbers of single base mismatch neighbors among the probes, a single base mismatch neighbor being another probe that has the same sequence except for a single base that is different;

computer readable code that sequences the nucleic acid sequence according to the selected nucleic acid probes; and

a tangible medium that stores the computer readable codes.

18. In a computer system, a method of sequencing a nucleic acid sequence, the method comprising the steps of:

inputting hybridization intensities for a plurality of nucleic acid probes;
selecting nucleic acid probes that have fewer than a predetermined number
of base mismatches with another probe; and
sequencing the nucleic acid sequence according to the selected nucleic
acid probes.

19. The method of claim 19, further comprising the step of selecting
nucleic acid probes with highest hybridization intensities.

20. The method of claim 19, wherein the nucleic acid probes with the
highest hybridization intensities are above an intensity threshold.

21. The method of claim 18, wherein each nucleic acid probe
includes n bases.

22. The method of claim 21, further comprising the steps of:
comparing bases at positions 1 through n of a first nucleic acid probe to
bases at positions 1 through n of a second nucleic acid probe; and
counting base mismatches between the first and second nucleic acid probe.

23. The method of claim 21, further comprising the steps of:
comparing bases at positions 1 through $n-1$ of a first nucleic acid probe to
bases at positions 2 through n of a second nucleic acid probe; and
counting base mismatches between the first and second nucleic acid probe.

24. The method of claim 18, wherein the sequencing step includes
the steps of:
aligning the selected nucleic acid probes so that bases that are common
overlap; and
sequencing the nucleic acid sequence according to bases that occur most
often at each position in the aligned nucleic acid probes.

25. The method of claim 24, wherein the nucleic acid sequence is sequenced as complementary to the bases that occur most often.

26. A computer program product that sequences a nucleic acid sequence, comprising:

computer readable code that receives as input hybridization intensities for a plurality of nucleic acid probes;

computer readable code that selects nucleic acid probes that have fewer than a predetermined number of base mismatches with another probe;

Computer readable code that sequences the nucleic acid sequence according to the selected nucleic acid probes; and

A tangible medium that stores the computer readable codes.

27. A method of sequencing a nucleic acid, comprising the steps of:
contacting a set of oligonucleotide probes of predetermined sequence and length with the nucleic acid under hybridization conditions that do not allow differentiation between (i) those probes of the set which are perfectly complementary to part of the nucleic acid and (ii) those probes that are not perfectly complementary to part of the nucleic acid;

selecting a subset of oligonucleotide probes that includes probes that are perfectly complementary to part of the nucleic acid and probes that are not perfectly complementary to part of the nucleic acid; and

determining the sequence of the nucleic acid by compiling overlapping sequences of the subset of probes.

28. The method of claim 27, further comprising the step of fragmenting the nucleic acid before contacting the nucleic acid with the set of oligonucleotide probes.

29. The method of claim 28, further comprising the step of labeling the nucleic acid before contacting the nucleic acid with the set of oligonucleotide probes.